Assignment 1

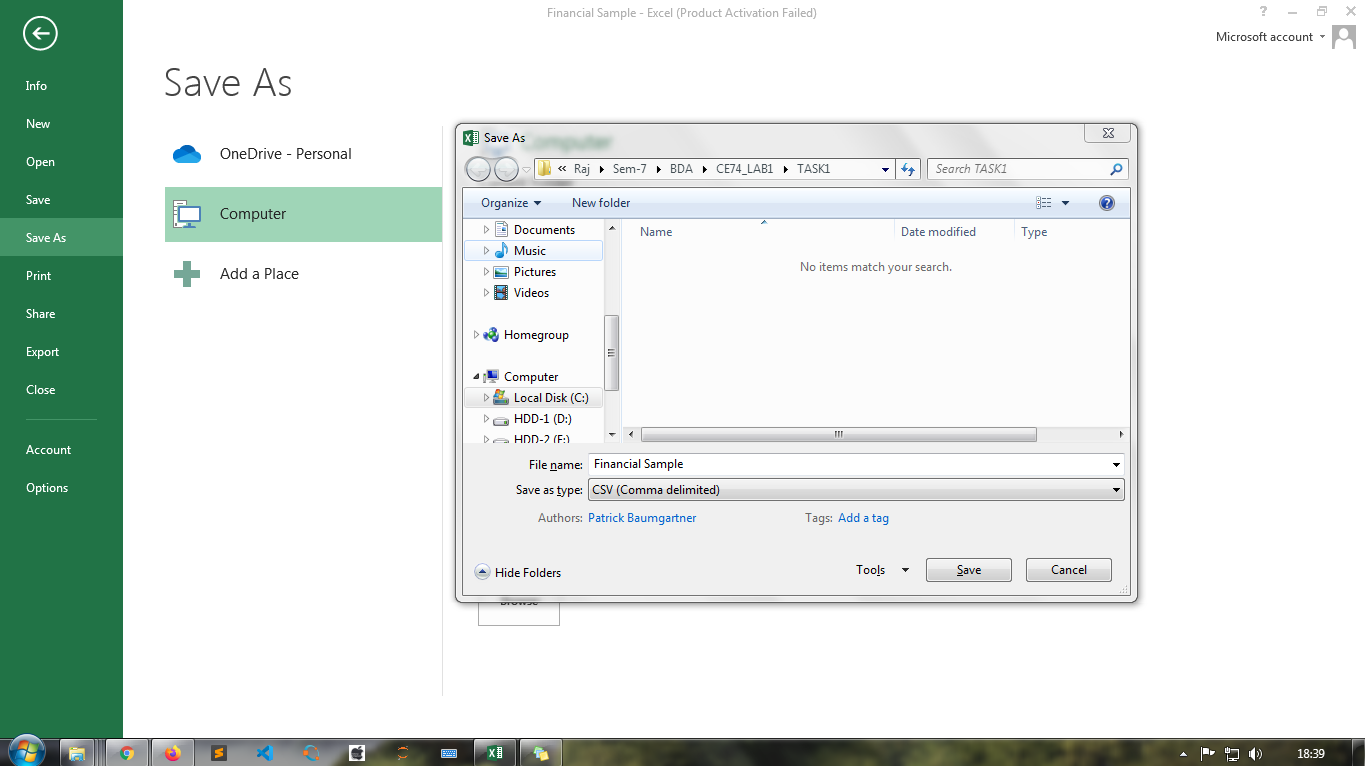
Objective: Recording types of data and various file formats. Identifying data sources. Handling traditionally to start with at a small scale.

Task 1: Given the spreadsheet file convert it into a csv.

There are two method that I have been known so far detailed below.

**Method 1: Using MS Excel.**

Click on the File button and then choose Save As. Choose CSV in Save as type and then press Save button.



**Method 2: Using Script.**

Write down below script in notepad and save as XlsToCsv.vbs in location where .xlsx file is stored. Open cmd for current location and execute following command.

if WScript.Arguments.Count < 2 Then

WScript.Echo "Please specify the source and the destination files. Usage: ExcelToCsv <xls/xlsx source file> <csv destination file>"

Wscript.Quit

End If

csv\_format = 6

Set objFSO = CreateObject("Scripting.FileSystemObject")

src\_file = objFSO.GetAbsolutePathName(Wscript.Arguments.Item(0))

dest\_file = objFSO.GetAbsolutePathName(WScript.Arguments.Item(1))

Dim oExcel

Set oExcel = CreateObject("Excel.Application")

Dim oBook

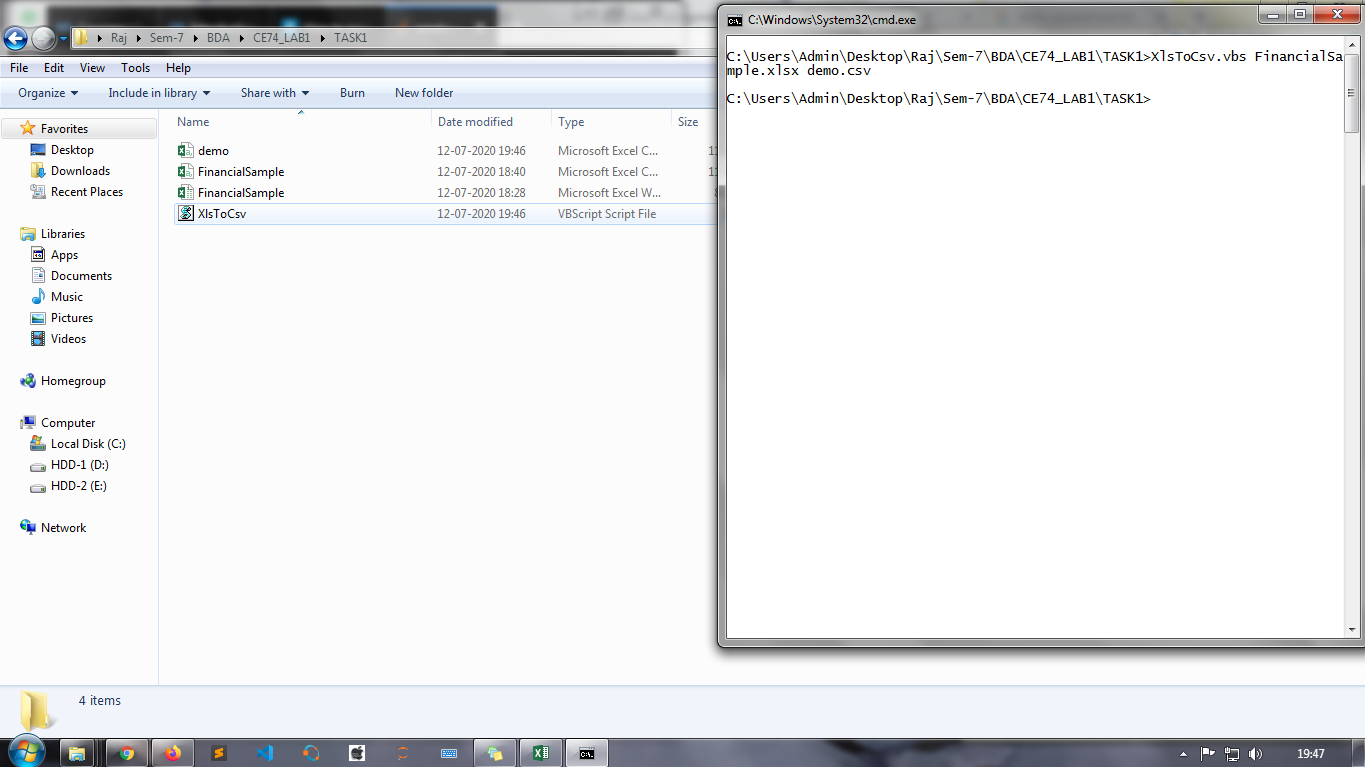
Set oBook = oExcel.Workbooks.Open(src\_file)

oBook.SaveAs dest\_file, csv\_format

oBook.Close False

oExcel.Quit

$ XlsToCsv.vbs [sourcefile].xlsx [destinationfile].csv



Task 2: Import a csv into MySQL database table.

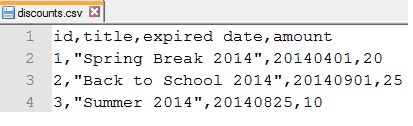
**Using LOAD DATA INFILE statement:**

LOAD DATA INFILE statement allows you to read data from a text file and import the file’s data into a database table very fast.

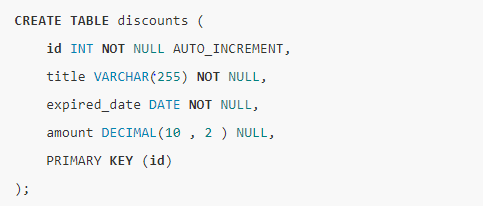
Before importing the file, you need to prepare the following:

* A database table to which the data from the file will be imported.
* A CSV file with data that matches with the number of columns of the table and the type of data in each column.
* The account, which connects to the MySQL database server, has FILE and INSERT privileges.

Suppose we have discount.csv file contains the first line as column headings and other three lines of data.



We use CREATE TABLE statement to create the discounts table as follows:



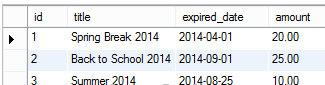
Finally, the following statement imports data from the [location].csv file into the discount table.



* The field of the file is terminated by a comma indicated by  FIELDS TERMINATED BY and enclosed by double quotation marks specified by ENCLOSED BY ‘ “ ’.
* Each line of the CSV file is terminated by a newline character indicated by LINES TERMINATED BY ’\n’.
* Because the file has the first line that contains the column headings, which should not be imported into the table, therefore we ignore it by specifying IGNORE 1 ROWS option.

Now, you can check the discount table to see whether the data is imported.

SELECT \* FROM discount;



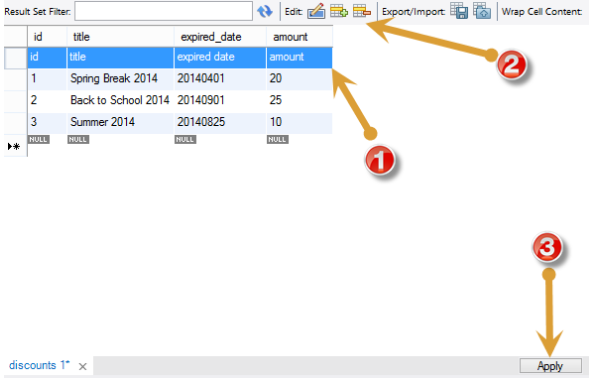
**Using MySQL Workbench:**

The following are steps that you want to import data into a table from MySQL workbench:

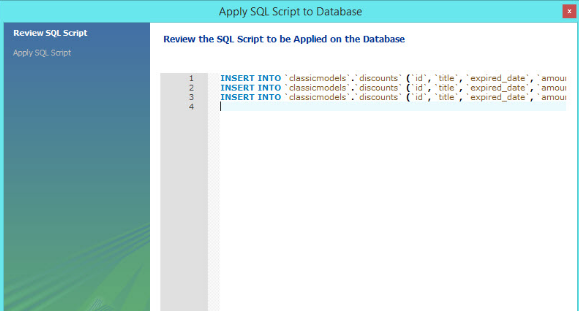
Open table to which the data is loaded.

Click import button. Choose a CSV file and click Open button.

Review the data, Click Apply button.



After, MySQL workbench will display a dialog “Apply SQL Script to Database”, click Apply button to insert data into the table.

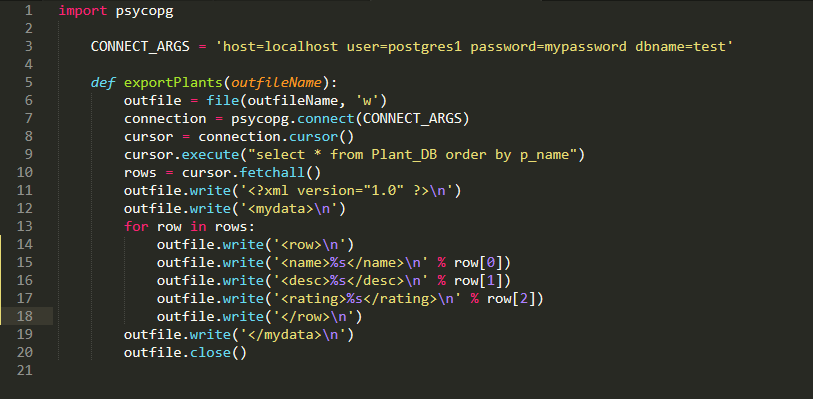


Task 3: Write a computer program to read records from database and generate data file.

**For XML :**

This task can perform by write a python application that reads your table , and writes data formatted as XML to a file. In python, below example reads a table in a postgreSQL database using psycopg- python postgreSQL database adapter.

Code :



Sample Output :



**For JSON :**

This example uses pyodbc, which support connections to SQL Server, MySQL, Microsoft Access and other databases. As shown in example script builds a list of dictionaries, with each row in the database becoming one dictionary and each field in the row a key-value pair.

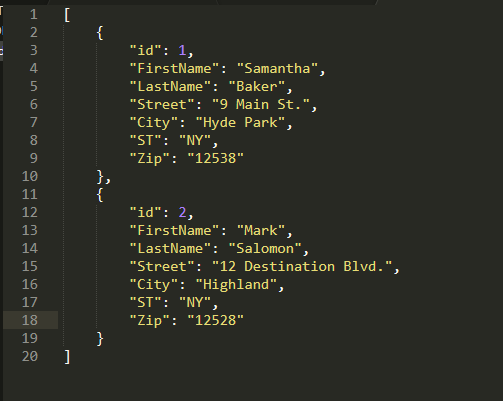
That’s it.

Code :



Output:

Student\_objects.js



Task 4: Import XML/JSON file into another database/table. I.e. MS Access. Oracle, etc.

**Import XML file into Oracle database table.**

XML file: employee.xml

<ROWSET>

<ROW ID="10">

<EmpName>Ritesh<EmpName>

<EmpSal>10000<EmpSal>

<DeptNo>1010<DeptNo>

<JoinDate>03-Apr-2006<JoinDate>

<ROW>

<ROW ID="20">

<EmpName>Arti<EmpName>

<EmpSal>20000<EmpSal>

<DeptNo>2020<DeptNo>

<JoinDate>05-May-2007<JoinDate>

<ROW>

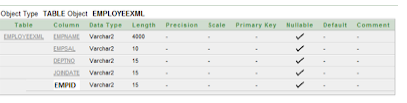
<ROWSET>

Solution :

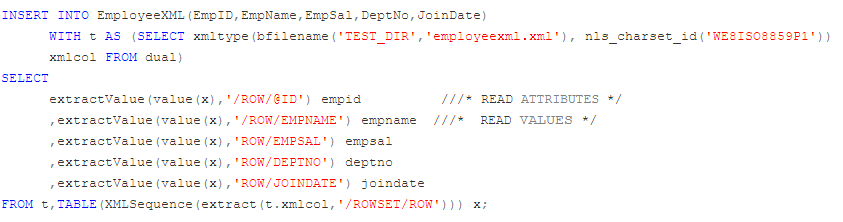
1. Create Virtual directory from oracle SQL Prompt.

SQL> CREATE directory test\_dir AS “c:\Test”;

1. Copy “employee.xml” file into “c:\Test”.
2. Create database Table name “EmployeeXML”.

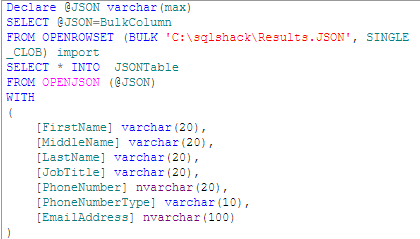


1. Oracle SQL query to load data into EmployeeXML Table.



**Import JSON file data into the SQL Server table**

We can use the similar t-SQL described earlier for inserting data in an existing table. We can also create a table using the SQL SELECT INTO statement similar to a relational DB table. The following query, imports data from results.json, creates a JSON table automatically using SQL SELECT INTO statement and inserts data into it:



We can verify the records using the following SELECT statement:

SELECT \* from JSONTable;

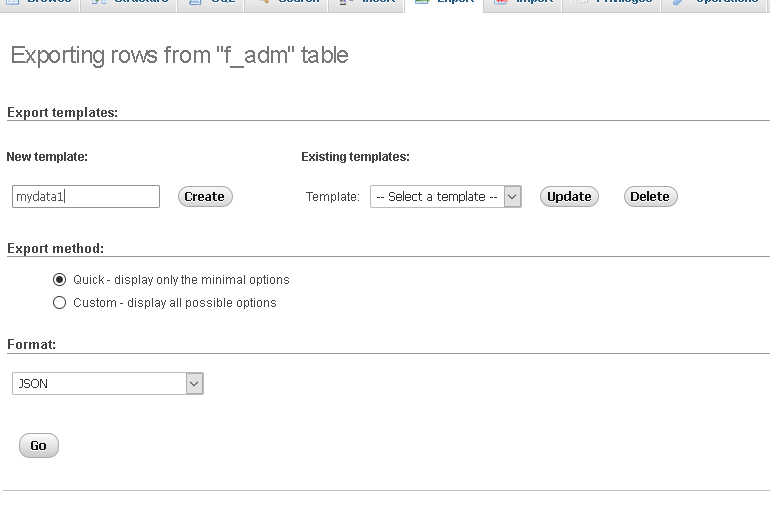
Task 5: Export database dump for data migration/archival.

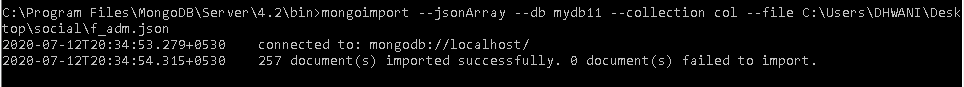
To migrate a database using Export/Import, complete the following three basic steps:

1. Export the data from the database you are migrating (the source database). The export physically copies the data to the export dump file.
2. Create the Oracle8*i* database into which you will import the exported data (the target database).
3. Import the exported data into the new Oracle8*i* target database

Task 6: Validate/Map data types across different database systems when migrating from one to another.

Export the JSON file from the phpmyAdmin.

****

Import the previously exported file in mongoDB with CLI.****

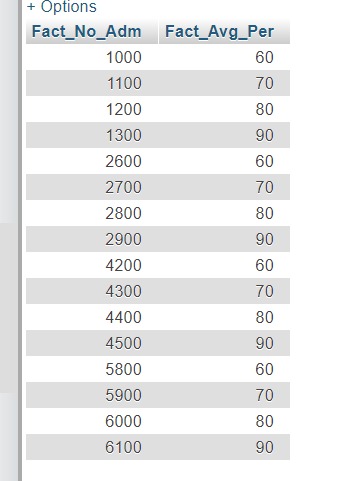
Task 7: Represent Data Cube and perform operation. OLAP – data warehouse

Data warehousing is the electronic storage of a large amount of information by a business or organization. A data warehouse is designed to run query and analysis on historical data derived from transactional sources for business intelligence and data mining purpose.

Slice :

Find and display number of students only those Male & Open type of fellow with their percentage average for each period and for each type of branch separately.

Select Fact\_No\_Adm, Fact\_Avg\_Per from JD\_Branch\_ADM B,JD\_Fellow\_ADM F,JD\_Period\_ADM D,JF\_ADM FACT where ( B.Branch\_Id = FACT.Branch\_Id and F.Fellow\_Id = FACT.Fellow\_Id and D.Period\_Id = FACT.Period\_Id and F.Fellow\_Gender = 'M' and F.Fellow\_Category = 'OPEN');

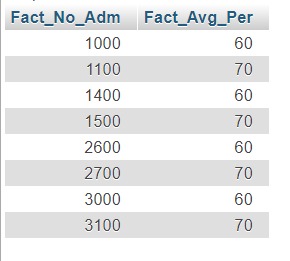


Dice :

Find number of students with their average percentage which satisfies following:

* Belong to Branch Category of Engineering
* Male
* Have reshuffled during year 2004

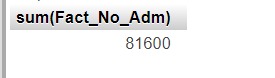
Select Fact\_No\_Adm, Fact\_Avg\_Per from JD\_Branch\_ADM B,JD\_Fellow\_ADM F,JD\_Period\_ADM D,JF\_ADM FACT where ( B.Branch\_Id = FACT.Branch\_Id and F.Fellow\_Id = FACT.Fellow\_Id and D.Period\_Id = FACT.Period\_Id and (F.Fellow\_Id = 1 or F.Fellow\_Id = 2) and (B.Branch\_Id = 1 or B.Branch\_Id = 2) and (D.Period\_Id = 1 or D.Period\_Id = 2));



Roll UP:

Find number of total students admitted to branch category of Engineering.

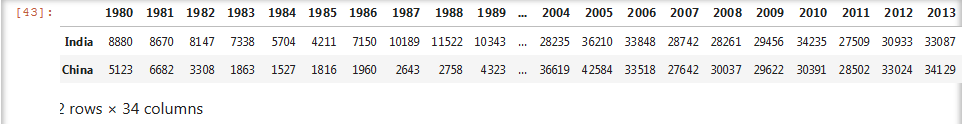
Select sum(Fact\_No\_Adm) from JD\_Branch\_ADM B,JD\_Fellow\_ADM F,JD\_Period\_ADM D,JF\_ADM FACT where B.Branch\_Id = FACT.Branch\_Id and F.Fellow\_Id = FACT.Fellow\_Id and D.Period\_Id = FACT.Period\_Id and B.Branch\_Category = 'ENG' group by B.Branch\_Category;

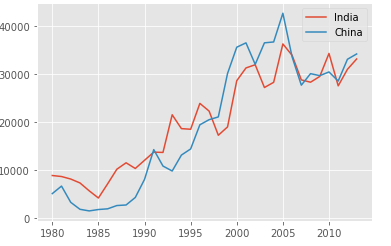


Task 8: Generate PDF report.

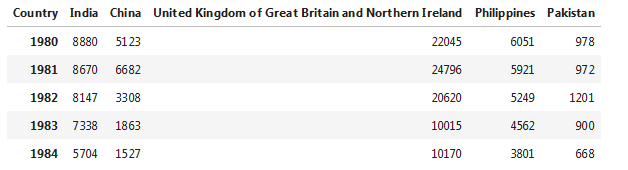
Dataset: Immigration to Canada from 1980 to 2013

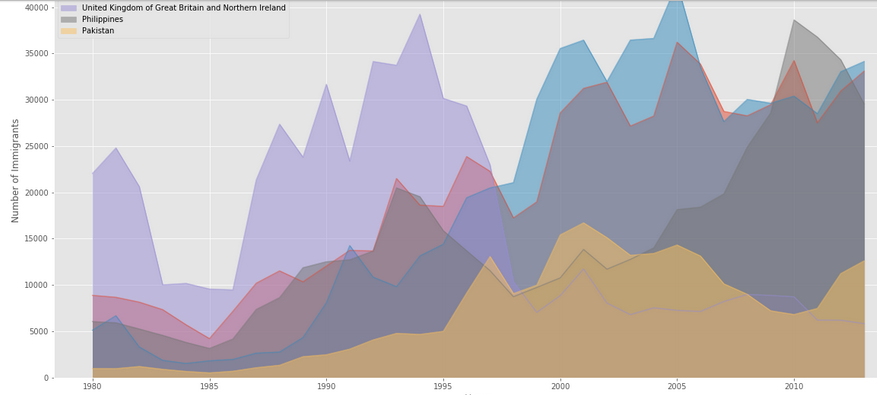
* Compare the number of immigrants from India and China from 1980 to 2013.



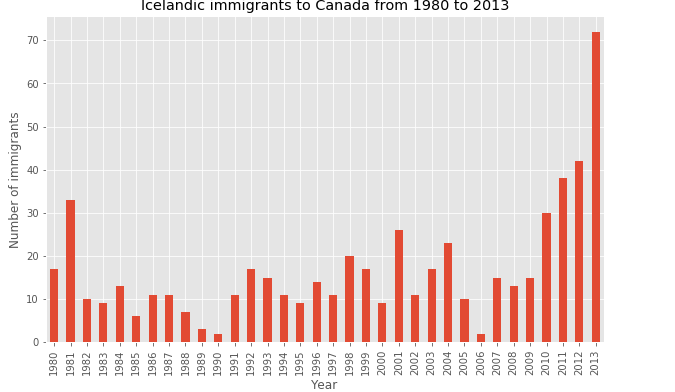


* Visualized the top 5 countries that contributed the most immigrants to Canada from 1980 to 2013.

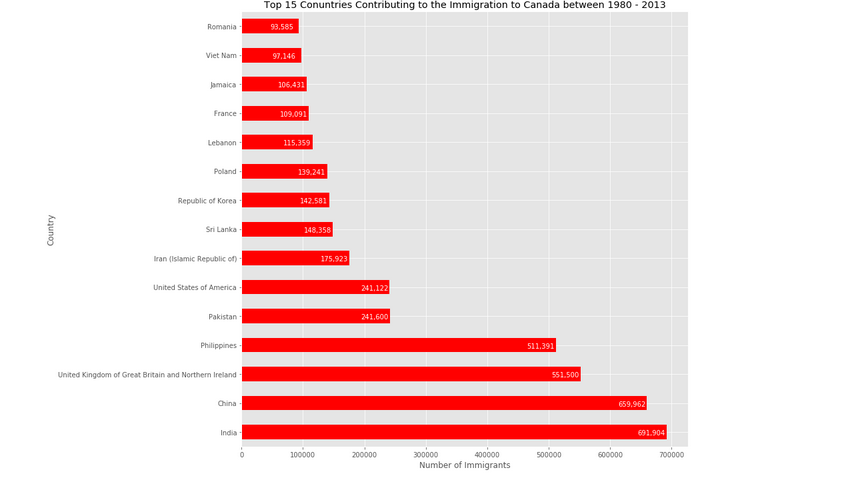




* Compare the number of Icelandic immigrants (country = 'Iceland') to Canada from year 1980 to 2013.



* Total number of immigrants to Canada from the top 15 countries, for the period 1980 – 2013.



* Proportion (percentage) of new immigrants grouped by continents for the entire time period from 1980 to 2013.

